DOCUMENT RESUME

ED 319 788 TM 015 041

AUTHOR	Marsh, Herbert W.; Hocevar, Dennis
TITLE	The Multidimensionality of Students' Evaluations of Teaching Effectiveness: The Generality of Factor Structures across Academic Discipline, Instructor Level, and Course Level.
PUB DATE	15 Jan 90
NOTE	18p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Curriculum; *College Faculty; *Factor Structure; *Graduate Students; Higher Education; Multidimensional Scaling; *Student Evaluation of Teacher Performance; *Teacher Effectiveness; Teaching Assistants; *Undergraduate Students
IDENTIFIERS	*Student Evaluation of Educational Quality; University of Southern California

ABSTRACT

Factor analyses of student evaluations of teaching effectiveness were conducted for 24,158 courses at the University of Southern California and for each of 21 different subgroups derived from the total group. All classes evaluated by six or more students were included in the study. The subgroups were designed to differ in terms of instructor level (teaching assistants or regular faculty), course level (undergraduate or graduate), and academic discipline. The same 9 factors that the Student Evaluation of Educational Quality instrument was designed to measure were consistently identified in each of the 22 different factor analyses, and all factor structures were remarkably well-defined and consistent. Correlations between factor scores based on the total group factor analysis and the 21 subgroup factor analyses were very high, and most were greater than 0.99. Because of the large number and diversity of classes in this study, the results provide stronger support for the generality of the factor structure underlying students' evaluations of teaching effectiveness than does any previous research. Five data tables are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED319788

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
HERBERT W. MARSH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Multidimensionality of Students' Evaluations of Teaching Effectiveness: The Generality of Factor Structures Across Academic Discipline, Instructor Level, and Course Level

Herbert W. Marsh

University of Sydney, NSW Australia

Dennis Hocevar

University of Southern California

15 January, 1990

Running Head: Multidimensionality of Students' Evaluations

TM 015 041

BEST COPY AVAILABLE

ABSTRACT

Factor analyses of student evaluations of teaching effectiveness were conducted for a total group 24,158 courses and for each of 21 different subgroups derived from the total group. The subgroups were constructed to differ in terms of instructor level (teaching assistants or regular faculty), course level (undergraduate or graduate), and academic discipline. The same nine factors that the Student Evaluation of Educational Quality (SEEQ) instrument was designed to measure were consistently identified in each of the 22 different factor analyses and all factor structures were remarkably well defined and consistent. Correlations between factor scores based on the total group factor analysis and the 21 subgroup factor analyses were very high and most very greater than .99. Because of the large number and diversity of classes in this study, the results provide stronger support for the generality of the factor structure underlying students' evaluations of teaching effectiveness than does any previous research.

Information from students' evaluations necessarily depends on the content of the evaluation items. Student ratings, like the teaching that they represent, should be viewed as a multidimensional construct (e.g., a teacher may be quite well organized but lack enthusiasm). This contention is supported by common-sense and a considerable body of empirical research. Unfortunately, most evaluation instruments and research fail to take cognizance of this multidimensionality. If a survey instrument contains an ill-defined hodge-podge of different items and student ratings are summarized by an average of these items, then there is no basis for knowing what is being measured, no basis for differentially weighting different components in a way that is most appropriate to the particular purpose they are to serve, nor any basis for comparing these results with other findings. If a survey contains separate groups of related items that are derived from a logical analysis of the content of effective teaching and the purposes that the ratings are to serve, or from a carefully constructed theory of teaching and learning, and if empirical procedures such as factor analysis demonstrate that the items within the same group do measure the same trait and that different traits are separate and distinguishable, then it is possible to interpret what is being measured.

In evaluating the need for multiple dimensions of students' evaluations it is important to consider the purposes that the ratings are intended to serve. Marsh (1984; 1987) noted that student ratings are used variously to provide and are recommended for purposes of: (a) formative feedback to faculty about the effectiveness of their teaching; (b) a summative measure of teaching effectiveness to be used in personnel decisions; (c) information for students to use in the selection of courses and instructors; (d) an outcome or a process description for research on teaching. Marsh (in press) argued that for 3 of these 4 recommended uses of students' evaluations — all but personnel decisions — there appears to be general agreement that appropriately constructed multiple dimensions of are more useful than a single summary score.

For personnel decisions, there is considerable debate as to whether a single score is more useful than of profile on scores reflecting multiple dimensions (see Abrami, 1988; Marsh, 1987; 1989). Some researchers, while accepting the multidimensionality of students' evaluations and the importance of measuring separate components for some purposes such as feedback to faculty, defend the unidimensionality of student ratings because "when student ratings are used in personnel decisions, one decision is made." There are, however, serious problems with this reasoning. First, the

use to which student ratings are put has nothing to do with their dimensionality, though it may influence the form in which the ratings are to be presented. Second, even if a single total score were the most useful form in which to summarize student ratings for personnel decisions -- and there is no reason to assume that generally it is -- this purpose would be poorly served by a ill-defined total score based upon an ad hoc collection of items that was not appropriately balanced with respect to the components of effective teaching that were being measured. If a single score were to be used, it should represent a weighted average of the different components where the weight assigned to each component was a function of logical and empirical analyses. Third, implicit in this argument is the suggestion that administrators are unable to utilize or prefer not to be given multiple sources of information for use in their deliberations, but I know of no empirical research to support such a claim.

The Content of the Multiple Dimensions

An important issue in the construction of multidimensional rating scale instruments is the content of the dimensions to be surveyed. The most typical approach consists of a logical analysis of the content of effective teaching and the purposes of students' evaluations, supplemented perhaps with literature reviews of the characteristics other researchers have found to be useful, and feedback from students and faculty. An alternative approach based on a theory of teaching or learning could be used to posit the important dimensions, though such an approach does not seem to have been used in student evaluation research. However, with each approach, it is important to also use empirical techniques such as factor analysis to further test the dimensionality of the ratings. The most carefully constructed instruments combine both logical/theoretical and empirical analyses in the research and development of student rating instruments.

The student evaluation literature does contain several examples of well constructed instruments with a clearly defined factor structure that provide measures of distinct components of teaching effectiveness. Some of these instruments and the factors that they measure are:

1) Frey's Endeavor instrument (Frey, Leonard & Beatty, 1975; also see Marsh, 1981, 1986): Presentation Clarity, Workload, Personal Attention, Class Discussion, Organization/Planning, Grading, and Student Accomplishments;

2) The Student Description of Teaching (SDT) questionnaire originally developed by Hildebrand, Wilson and Dienst (1971): Analytic/Synthetic Approach, Organization/Clarity, Instructor Group Interaction, Instructor Individual Interaction, and Dynamism/Enthusiasm;

3) Marsh's Student Evaluations of Educational Quality (SEEQ) instrument (Marsh, 1982b; 1983, 1984; 1987; also see Table 3 presented later): Learning/Value, Instructor Enthusiasm, Organization/Clarity, Individual Rapport, Group Interaction, Breadth of Coverage, Examinations/Grading, Assignments/Readings, and Workload/Difficulty;

4) The Michigan State SIRS instrument (Warrington, 1973): Instructor Involvement, Student Interest and Performance, Student-Instructor Interaction, Course Demands, and Course Organization.

The systematic approach used in the development of these instruments and the similarity of the factors which they measure, supports their construct validity. Factor analyses of responses to each of these instruments provide clear support for the factor structure they were designed to measure, and demonstrate that the students' evaluations do measure distinct components of teaching effectiveness. More extensive reviews describing the components found in other research (Cohen, 1981; Feldman, 1976; Kulik & McKeachie, 1975) identify dimensions similar to those described here.

Factor analysis is a useful technique for determining what factors are being measured, but it cannot determine whether the obtained factors are important to the understanding of effective teaching. Consequently, carefully developed surveys — even when factor analysis is to be used — typically begin with item pools based upon literature reviews, and with systematic feedback from students, faculty, and administrators about what items are important and what type of feedback is useful (e.g., Hildebrand, Wilson & Dienst, 1971; Marsh, 1982b). For example, in the development of SEEQ a large item pool was obtained from a literature review, instruments in current usage, and interviews with faculty and students about characteristics which they see as constituting effective teaching. Then, students and faculty were asked to rate the importance of items, faculty were asked to judge the potential usefulness of the items as a basis for feedback, and open-ended student comments on pilot instruments were examined to determine if important aspects had been excluded. These criteria, along with psychometric properties, were used to select items and revise subsequent versions. This systematic development constitutes evidence for the content validity of SEEQ and makes it unlikely that it contains any irrelevant factors.

Feldman (1976; also see Feldman, 1983, 1984, 1986, 1987) categorized the different characteristics of the superior university teacher from the student's point of view with a systematic review of research that either asked students to specify these characteristics or inferred them on the

basis of correlations between specific characteristics and students' overall evaluations. On the basis of such studies, and also to facilitate presentation of this material and his subsequent reviews of other student evaluation research, Feldman derived a set of categories shown in Table 1. This list provides the most extensive and, perhaps, the best set of characteristics that are likely to underlie students' evaluations of effective teaching. Nevertheless, Feldman used primarily a logical analysis based on his examination of the student evaluation literature, and his results do not necessarily imply that students can differentiate these characteristics. Also, to actually measure all these characteristics as separate scales would require an instrument that contained as many as 100 items and this would be unacceptable in most settings. This set of characteristics does, however, provide a useful basis for evaluating the comprehensiveness of the set of evaluation factors on a given instrument.

Insert Table 1 About Here

Feldman (1976) noted that factors actually identified by factor analysis typically correspond to more than one of his categories. The highest loading items on any given factor often come from more than one of his categories. In Table 1 I have attempted to match Feldman's categories to the SEEQ factors. The only categories that are apparently unrelated to any of the SEEQ factors are teacher elocution (category 7) and, perhaps, teacher sensitivity to class level and progress (category 9). All of the SEEQ factors represent at least one of Feldman's categories and most reflect two or more categories. In contrast, none of Feldman's categories reflect more than one of the SEEQ factors. This logical analysis of the content of the SEEQ factors and Feldman's categories demonstrates that there is substantial overlap between the two but that Feldman's categories reflect more narrowly defined constructs than do the SEEQ factors.

The Present Investigation

The purpose of the present investigation is to extend previous work on the factor structure of SEEQ responses. Factor analysis provides tests of whether: (a) students are able to differentiate among components of effective teaching, (b) the empirical factors confirm the factors that an instrument is designed to measure, and (c) the same factors are identified consistently in different settings and in different academic disciplines. Factor scores derived from factor analysis also provide potentially useful scores for summarizing the results of students' evaluations.

SEEQ has been used across a diverse array of academic disciplines at the University of Southern California since 1976. For present purposes ratings from 24,158 different classes were divided into 21 different groups varying

in terms of instructor level (teaching assistant or regular academic staff), the course level (undergraduate or graduate), and the academic discipline. Twenty-two separate factor analyses were conducted; one for the total group and one for each of the 21 different subgroups. The factor structure of SEEQ responses was evaluated by comparing the results of these factor analyses and comparing empirically derived factor scores based on these factor analyses. The availability of such a large data base using on a psychometrically sound instrument provides much stronger tests of the comparability of the factor structures across instructor level, course level, and academic discipline than any previous research.

Methods

Sample and Procedures.

During the period 1976-1988, SEEQ forms were administered in approximately 40,000 courses at the University of Southern California. Although the use of the SEEQ form is voluntary, the University requires that each academic unit collect some form of students' evaluations of teaching effectiveness for all courses and staff are not considered for promotion unless students' evaluations are provided. Most of the academic units that do use SEEQ require that all their staff are evaluated in all courses. The evaluation forms are typically distributed to staff shortly before the end of each academic term, administered and collected by a student in the class or by a member of the academic staff according to printed instructions, and taken to a central office where they are processed. This program, the SEEQ instrument on which it is based, and research that led to its development are described by Marsh (1987).

For present purposes all classes taught by regular academic staff or by teaching assistants that were evaluated by 6 or more students were considered. Excluded were classes evaluated by 5 or fewer students and classes taught by teachers who were not graduate student teaching assistants or regular academic staff (i.e., had academic titles other than assistant, associate, or full professor). This resulted in a sample of 24,158 classes. Each of these classes was then classified into 21 subsamples such that each subsample had at least 400 classes. All classes were first categorized into three general groups consisting of classes taught by teaching assistants, undergraduate classes taught by regular faculty, and graduate courses taught by regular faculty. Classes were then classified into divisions or schools (e.g., Social sciences, Engineering) and then into specific departments (e.g., psychology, systems engineering) whenever there were more than 400 classes. Thus, each class was assigned to the most specific subgroup for which there

was at least 400 classes. All classes were classified into one and only one subgroup. This procedure resulted in 7 groups of classes taught by teaching assistants, 7 groups of undergraduate classes taught by regular faculty, and 7 groups of graduate classes taught by regular faculty (see Table 2).

Insert Table 2 About Here

Statistical Analyses

All analyses were performed on class-average responses for the total sample and for each of the 21 subsamples. The factor analyses of the 35 SEEQ items consisted of principal axis factoring with a Kaiser normalization and iterations followed by an oblique rotation (see SPSS, 1986). For each factor analysis empirically defined factor scores were generated using the regression method (SPSS, 1986).

Everett and Entrekin (1980; Everett, 1983) noted that factors extracted from responses to the same items administered to different samples should be comparable if they are to be used as summary measures. They went on to argue that correlations between factor scores derived from two different factor analyses "provides a coefficient of factor comparability, which is a more direct measure than the coefficient of congruence based upon factor loadings" (p. 165). Everett (1983) further demonstrated that this procedure provided a useful indication of the number of factors that should be retained. When factor scores based on a large number of different groups are considered, Marsh (1988) proposed a variation of this procedure that is used in the present investigation. In this variation, factor scores based on the total group and those based on each separate group are compared. Thus, for each case there is a set of factor scores based on the total group and a set of factor scores based on the particular group to which that case belongs. Correlations between matching factors in the two sets of factor scores provide an index the factor comparability between the factor analysis based on the total sample and that based on each subsample.

Results

The factor analysis based on the total sample of 24,258 classes is summarized in Tables 3 and 4. The factor analysis clearly identifies each of the 9 factors that SEEQ is designed to measure and the factor structure is very well defined. The 35 target loadings -- factor loadings for items designed to measure each factor (presented in boxes in Table 3) -- are consistently large; the mean is .650 and every one is at least .392. The 280 nontarget loadings are consistently small; the mean is .067 and none is larger than .245. Whereas the factor pattern correlations indicate that the factors are positively correlated (mean r = .318), the largest correlation is .512. These results replicate previous research with SEEQ.

Insert Tables 3 and 4 About Here

The set of 21 factor analyses conducted on each subsample is summarized in Table 4. For each of the 21 factor analyses the target loadings are consistently high (means of .578 to .712), nontarget loadings are consistently low (means of .062 to .076), and factor correlations are moderate (means of .257 to .399). The results are similar in the 7 sets of courses taught by teaching assistants, the 7 sets of undergraduate courses taught by regular faculty, and the 7 sets of graduate courses taught by regular faculty. The similarity of the factor structures across the 22 different factor analyses provides remarkably strong support for the generality of SEEQ factor structure and much stronger support than has been demonstrated with any other student evaluation instruments.

Although not emphasized in Table 4, there is also a consistent pattern for the nontarget loadings in the 22 factor analyses. The overall instructor rating and the overall course rating are most strongly related to the Instructor Enthusiasm and Learning/Values factors respectively. Not surprisingly, however, these overall rating items typically have moderate factor loadings on several other factors. For example, in the total group factor analysis (Table 3) the largest two nontarget loadings are for the overall instructor item on the Organization/Clarity factor and the for the overall course item on the Instructor Enthusiasm factor. Across all the factor analyses one of the overall rating items had the largest nontarget loadings in 13 of the 22 analyses. These results demonstrate that particularly the overall rating items are not pure indicators of any of the factors.

Inspection of Table 3 indicates that most of the nontarget loadings are close to zero or modestly positive. However, nearly half nontarget loadings for the Workload/Difficulty items on the remaining SEEQ factors and the nontarget loadings for the remaining 31 items on the Workload/Difficulty factor are negative. Similarly, the lowest correlations among SEEQ factors consistently involve the Workload/Difficulty factor. These patterns observed for the total group factor analysis are found consistently in the 21 subgroup factor analyses.

Two sets of factor scores representing the nine SEEQ factors were generated for each of the 24,158 classes. One set of factor scores was based on the factor analysis of the total sample presented in Table 3 and the second set of factor scores was based on the particular subsample to which the class belonged. These two sets of factor scores were then correlated for each of the 21 different subsamples. High correlations between factor scores representing the same factor provide support for the comparability of the

different factor structures. Nearly all of these 189 correlations are greater than .95 and a majority are larger than .99 (Table 5). Low correlations between factor scores representing different factors provide support for the differentiation among the factors. The means correlations among these nonmatching factor scores vary from .254 to .446 for the 21 subsamples.

Insert Table 5 About Here

Discussion

Research reviewed here suggests that most -- if not all -- of the recommended purposes of students' evaluations of teaching effectiveness are better served by an appropriately constructed set of multiple dimensions than by a single summary score. Most evaluation instruments, however, do not measure a well-defined set of evaluation factors and most research does not incorporate this multidimensional perspective. An important exception to this generalization is the SEEQ instrument and research based upon it that are the focus of the present investigation. Four observations support the appropriateness of the factors used to summarize SEEQ responses. First, empirical research supports the SEEQ factor structure. Second, the systematic development of the SEEQ instrument supports the content validity of its factors. Third, factor analyses of other instruments designed to measure multiple dimensions of teaching effectiveness result in factors like those identified by SEEQ responses. Finally, there is substantial agreement between the content of empirically derived SEEQ factors and the categories of effective teaching developed by Feldman (1976).

The purpose of the present investigation was to extend previous research by evaluating the generality of the SEEQ factor structure across a very large and diverse set of different classes. The SEEQ factor structure was well defined for the total sample of 24,158 classes and this result replicates previous research based on smaller samples. Of particular importance, was the finding that the factor structure was also well defined for 21 subgroups that varied in terms of instructor level, course level, and academic discipline. The nine factors that SEEQ is designed to measure were identified in all 22 factor analyses and factor scores based on the total group analysis were almost perfectly correlated with those based on each of the 21 subgroup analyses. Because of the psychometric properties of the SEEQ instrument and because of the size and diversity of the data base considered here, the results provide much stronger support for the generality of the factor structure underlying students' evaluations of teaching effectiveness than does any previous research.

REFERENCES

Abrami, P. C. (1989). SEEQint the truth about student ratings of instruction. Educational researcher, 81, 43-45.

Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. Review of Educational Research, 51, 281-309.

Everett, J. E. (1983). Factor comparability as a means of determining the number of factors and their rotation. Multivariate Behavioral Research, 18 197-218.

Everett, J. E., Entrekin, L. V. (1980). Factor comparability and the advantages of multiple group factor analysis. Multivariate Behavioral Research, 2 165-180.

Feldman, K. A. (1976). The superior college teacher from the student's view. Research in Higher Education, 5,243-288.

Feldman, K. A. (1983). The seniority and instructional experience of college teachers as related to the evaluations they receive from their students. Research in Higher Education, 18, 3-124.

Feldman, K. A. (1984). Class size and students' evaluations of college teacher and courses: A closer look. Research in Higher Education, 21, 45-116.

Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics: A review and synthesis. Research in Higher Education, 24, 139-213.

Feldman, K. A. (1987). Research productivity and scholarly accomplishment: A review and exploration. Research in Higher Education, 227-298.

Frey, P. W., Leonard, D. W., & Beatty, W. W. (1975). Student ratings of instruction: Validation research. American Educational Research Journal, 12, 327-336.

Hildebrand, M., Wilson, R. C., & Dienst, E. R. (1971). Evaluating university teaching. Berkeley: Center for Research and Development in Higher Education, University of California, Berkeley.

Kulik, J. A., & McKeachie, W. J. (1975). The evaluation of teachers in higher education. In Kerlinger (Ed.), Review of Research in Education, (Vol. 3). Itasca, IL: Peacock.

Marsh, H. W. (1981). Students' evaluations of tertiary instruction: Testing the applicability of American surveys in an Australian setting. Australian Journal of Education, 25, 177-192.

Marsh, H. W. (1982b). SEEQ: A reliable, valid, and useful instrument for collecting students' evaluations of university teaching. British Journal of Educational Psychology, 52, 77-95.

Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility. Journal of Educational Psychology, 76, 707-754.

Marsh, H. W. (1986). Applicability paradigm: Students' evaluations of teaching effectiveness in different countries. Journal of Educational Psychology, 78, 465-473.

Marsh, H. W. (1987). Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research. International Journal of Educational Research, 11, 253-388. (Whole Issue No. 3)

Marsh, H. W. (1989). Responses to reviews of "Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research" Instructional Evaluation, 10, 5-9.

SPSS (1986). SPSSx User's Guide. New York: McGraw-Hill.

Warrington, W. G. (1973). Student evaluation of instruction at Michigan State University. In A. L. Sockloff (Ed.), Proceedings: The first invitational conference on faculty effectiveness as evaluated by students (pp. 164-182). Philadelphia: Measurement and Research Center, Temple University.

Table 1

Categories of Effective Teaching Adapted From Feldman (1976, 1983, 1984 1986, 1987) and the SEEQ factors Most Closely Related to Each Category

Feldman's Categories	SEEQ Factors
1) Teacher's stimulation of interest in the course and subject matter.	Instructor Enthusiasm
2) Teacher's enthusiasm for subject or for teaching.	Instructor Enthusiasm
3) Teacher's knowledge of the subject.	Breadth of Coverage
4) Teacher's intellectual expansiveness and breadth of coverage.	Breadth of Coverage
5) Teacher's preparation and organization of the course.	Organization/Clarity
6) Clarity and understandableness of presentations and explanations.	Organization/Clarity
7) Teacher's elocutionary skills.	None
8) Teacher's sensitivity to, and concern with, class level and progress.	None
9) Clarity of course objectives and requirements.	Organization/Clarity
10) Nature and value of the course material including its usefulness and relevance.	Assignments/Readings
11) Nature and usefulness of supplementary materials and teaching aids.	Assignments/Readings
12) Perceived outcome or impact of instruction.	Learning/Value
13) Teacher's fairness and impartiality of evaluation of students; quality of exams.	Examinations/Grading
14) Personal Characteristics (personality)	Instructor Enthusiasm
15) Nature, quality and frequency of feedback from teacher to students.	Examinations/Grading
16) Teacher's encouragement of questions and discussion, and openness to the opinions of others.	Group Interaction
17) Intellectual challenge and encouragement of independent thought.	Learning/Value
18) Teacher's concern and respect for students; friendliness of of the teacher.	Individual Rapport
19) Teacher's availability and helpfulness.	Individual Rapport
20) Difficulty and workload of the course.	Workload/Difficulty

Note. The actual categories used by Feldman in different studies (e.g., Feldman, 1976, 1983, 1984, 1986, 1987) varied somewhat. Categories 12 and 14 were not included in Feldman (1976) but were included in subsequent studies whereas category 20 was included by Feldman (1976) but not subsequent studies. One other category (classroom management) was only included by Feldman (1976).

[a] Whereas these SEEQ factors most closely match the corresponding categories, the match is apparently not particularly close.

Table 2

Summary of the 21 Subsamples of Courses

	N of Classes	Academic Unit
Undergraduate Courses Taught By Teaching Assistants		
1	431	General
2	610	Business
3	565	Humanities
4	1606	Social Sciences
5	683	Spanish and Portuguese
6	1368	Economics
7	902	Communication
Undergraduate Courses Taught By Regular Faculty		
1	1421	General
2	2326	Business
3	956	Humanities
4	2320	Social Sciences
5	1693	Engineering
6	590	History
7	538	Psychology
Graduate Courses Taught By Regular Faculty		
1	757	General
2	2049	Business
3	1157	Social Sciences
4	957	Engineering
5	1213	Education
6	457	Systems Engineering
7	1559	Safety and Systems Management
Total		
	24158	

Note. For present purposes all classes with 6 or more sets of ratings were classified into 21 subsamples such that each subsample had at least 400 classes. All classes were first categorized into general groups consisting of classes taught by teaching assistants, undergraduate classes taught by regular faculty, and graduate courses taught by regular faculty. Classes were then classified into divisions or schools (e.g., Social sciences or Engineering) and then into specific departments (e.g., psychology or systems engineering) whenever there were more than 400 classes. All classes were classified into one and only one subsample.

Table 3

Factor Analysis Results For the Total Sample of 24,158 Sets of Class-average Responses: Factor Loadings and Factor Correlations

SEEQ Scales and Items (paraphrased)	SEEQ Factors								
	Lrn	Enth	Orgn	Grp	Ind	Brd	Exam	Asgn	Work
Learning/Value									
Course challenging & stimulating	.434	.168	.103	.015	.014	.159	.099	.155	.291
Learned something valuable	.607	.083	.100	.026	.050	.103	.085	.147	.113
Increase subject interest	.646	.078	.034	.039	.058	.169	.074	.131	.020
Learned & understood subject matter	.487	.043	.176	.152	.045	.047	.112	.149	-.217
Overall Course Rating	.410	.211	.173	.041	.042	.085	.166	.175	.069
Instructor Enthusiasm									
Enthusiastic about teaching	.095	.544	.129	.072	.195	.115	.052	.069	.025
Dynamic and energetic	.064	.714	.094	.039	.085	.083	.069	.071	.042
Enhanced presentation with humor	.089	.650	-.023	.103	.078	.129	.090	.054	-.045
Teaching style held your interest	.137	.581	.187	.131	.026	.050	.110	.073	.017
Overall Instructor Rating	.172	.392	.245	.083	.141	.096	.140	.075	.039
Organization/Clarity									
Lecturer explanations clear	.146	.165	.510	.176	.060	.075	.079	.104	-.072
Materials well explained & prepared	.069	.087	.677	.060	.075	.073	.094	.118	.005
Course objectives stated & pursued	.128	.026	.529	.055	.070	.065	.175	.184	.024
Lectures facilitated taking notes	.031	.040	.589	-.093	.049	.175	.146	.044	.020
Group Interaction									
Encouraged class discussion	.038	.103	.011	.769	.070	.033	.067	.080	.002
Students shared knowledge/ideas	.066	.049	-.015	.797	.095	.093	.048	.073	-.029
Encouraged questions & gave answers	.059	.105	.167	.583	.151	.094	.100	.080	.001
Encouraged expression of ideas	.045	.069	.035	.674	.182	.110	.094	.070	-.013
Individual Rapport									
Friendly towards individual students	.051	.163	-.001	.176	.612	.063	.112	.057	-.038
Welcomed students seeking help/advice	.042	.059	.061	.078	.786	.036	.093	.059	-.007
Interested in individual students	.086	.140	.001	.137	.647	.057	.138	.059	.004
Accessible to individual students	-.014	-.028	.139	.037	.636	.099	.136	.104	.010
Breadth of coverage									
Contrasted various implications	.043	.037	.118	.059	.068	.676	.077	.109	.065
Gave background of ideas/concepts	.087	.085	.134	.020	.044	.662	.056	.122	.004
Gave different points of view	.035	.066	.086	.123	.101	.636	.097	.113	-.004
Discussed current developments	.207	.113	.018	.086	.039	.562	.084	.040	.000
Examinations/Grading									
Examination feedback valuable	.034	.039	.111	.047	.101	.028	.670	.088	.044
Evaluation methods fair/appropriate	.047	.044	.011	.043	.107	.078	.749	.099	-.033
Tested course content as emphasized	.063	.036	.129	.034	.064	.047	.643	.146	-.026
Assignments/Readings									
Readings/texts were valuable	-.008	-.004	.019	.022	.018	.053	.025	.885	-.003
They contributed to understanding	.127	.021	.036	.027	.039	.012	.140	.716	.072
Workload/Difficulty									
Course difficulty (easy-hard)	-.028	.030	.051	-.059	-.017	.096	.015	.018	.861
Course workload (light-heavy)	.100	-.054	.004	.085	-.001	.002	-.035	.038	.907
Course pace (slow-fast)	-.098	.101	.055	-.099	.005	-.001	.035	.040	.689
Hours per week outside of class	.148	-.044	-.085	.034	-.001	-.006	-.006	.042	.798

Factor Pattern Correlations	Lrn	Enth	Orgn	Grp	Ind	Brd	Exam	Asgn	Work
Learning/Value	1.000								
Instructor Enthusiasm	.434	1.000							
Organization/Clarity	.407	.427	1.000						
Group Interaction	.350	.364	.210	1.000					
Individual Rapport	.263	.400	.331	.455	1.000				
Breadth of Coverage	.449	.419	.454	.327	.352	1.000			
Examinations/Grading	.401	.392	.511	.315	.493	.403	1.000		
Assignments/Readings	.488	.319	.431	.312	.338	.418	.510	1.000	
Workload/Difficulty	.128	.076	.044	-.072	-.009	.106	.033	.154	1.000

Note. Target loadings, the factor loadings items designed to define each SEEQ factor, are presented in boxes.

Table 4

Summary of Factor Analyses Conducted on the Total Sample and on Each of the 21 Subsamples: Target Loadings, Nontarget Loadings, and Factor Correlations

	Target Loadings			Nontarget Loadings			Factor Correlations			N of
	Max	Min	Mean	Max	Min	Mean	Max	Min	Mean	Classes
Undergraduate Courses Taught By Teaching Assistants										
1	.940	.364	.578	.224a	-.142	.076	.575	.012	.388	431
2	.925	.415	.667	.246a	-.220	.066	.486	-.049	.298	610
3	.966	.332	.597	.278	-.312	.072	.561	-.190	.304	565
4	.900	.399	.662	.246a	-.195	.063	.515	-.152	.300	1606
5	.871	.285	.622	.261	-.276	.067	.522	-.059	.324	683
6	.893	.351	.620	.272	-.291	.068	.517	-.025	.333	1368
7	.888	.365	.630	.275a	-.227	.069	.538	-.062	.334	902
Mn	.912	.359	.625	.257	-.238	.068	.531	-.075	.326	
Undergraduate Courses Taught By Faculty										
1	.918	.414	.650	.224a	-.166	.066	.538	.011	.312	1421
2	.881	.408	.662	.217a	-.240	.069	.516	-.117	.317	2326
3	.892	.364	.633	.236	-.232	.071	.548	.057	.342	956
4	.881	.366	.653	.207	-.215	.068	.513	-.272	.306	2320
5	.933	.336	.642	.213a	-.226	.067	.509	.008	.314	1693
6	.925	.385	.666	.292a	-.300	.063	.508	-.010	.264	590
7	.947	.315	.617	.326	-.276	.072	.505	-.099	.313	538
Mn	.911	.370	.646	.245	-.236	.068	.520	-.060	.310	
Graduate Courses Taught By Faculty										
1	.959	.337	.640	.248a	-.142	.067	.530	-.079	.303	757
2	.924	.367	.672	.231a	-.206	.067	.527	-.092	.292	2049
3	.927	.348	.642	.256a	-.176	.067	.568	-.009	.313	1157
4	.829	.303	.605	.412	-.209	.067	.552	-.172	.270	957
5	.885	.359	.643	.271	-.084	.074	.511	-.092	.304	1213
6	.933	.438	.712	.251	-.154	.062	.469	-.128	.257	457
7	.947	.335	.623	.302a	-.283	.070	.543	-.114	.324	1559
Mn	.915	.355	.648	.282	-.189	.068	.528	-.098	.295	
Total Sample										
	.907	.392	.650	.245a	-.218	.067	.512	-.073	.318	24158

Note. Factor analyses were conducted on the total sample (see Table 1) and each of 21 subsamples. Factor loadings and factor correlations from these 22 factor analyses are summarized here.

[a] The largest nontarget loading was for either the overall teacher rating or the overall course rating.

Table 5

Correlations Between Factor Scores Based on the Total Sample and Based on the 21 Individual Samples

	Student Evaluation Factors									Mean Match	Mean Non-Match
	Lrn	Enth	Orgn	Grp	Ind	Brd	Exam	Asgn	Work		
Undergraduate Courses Taught By Teaching Assistants											
1	.995	.991	.986	.994	.987	.989	.988	.992	.996	.991	.446
2	.996	.990	.990	.993	.997	.997	.988	.996	.996	.994	.310
3	.799	.973	.979	.975	.993	.990	.990	.903	.986	.954	.338
4	.990	.997	.993	.995	.994	.994	.996	.997	.996	.995	.314
5	.984	.994	.990	.989	.998	.993	.983	.992	.993	.991	.349
6	.980	.989	.995	.998	.993	.998	.995	.995	.995	.993	.368
7	.980	.995	.995	.995	.997	.994	.991	.998	.991	.993	.360
Mn	.960	.990	.990	.991	.994	.993	.990	.982	.993	.987	.355
Undergraduate Courses Taught By Regular Faculty											
1	.993	.996	.990	.999	.998	.997	.997	.993	.992	.995	.318
2	.997	.996	.996	.997	.999	.996	.999	.991	.996	.996	.341
3	.993	.996	.991	.997	.997	.995	.996	.995	.994	.995	.379
4	.996	.999	.998	.999	.999	.996	.999	.999	.996	.998	.350
5	.996	.994	.992	.997	.998	.995	.997	.998	.992	.995	.328
6	.982	.980	.972	.995	.997	.984	.991	.946	.995	.982	.272
7	.913	.990	.956	.980	.990	.990	.994	.953	.995	.973	.359
Mn	.981	.993	.985	.995	.997	.993	.996	.982	.994	.991	.335
Graduate Courses Taught By Regular Faculty											
1	.987	.991	.965	.997	.991	.991	.986	.994	.995	.989	.322
2	.990	.996	.992	.998	.998	.991	.998	.993	.993	.994	.311
3	.976	.994	.977	.996	.995	.987	.998	.989	.997	.990	.329
4	.948	.983	.992	.994	.996	.982	.975	.773	.917	.951	.323
5	.979	.996	.991	.989	.992	.994	.997	.987	.996	.991	.318
6	.994	.994	.994	.993	.994	.983	.995	.994	.994	.993	.254
7	.989	.981	.983	.995	.995	.993	.993	.999	.996	.991	.354
Mn	.980	.991	.985	.995	.994	.989	.992	.961	.984	.986	.316

Note. Factor scores were generated from the factor analyses of the total sample and each of 21 subsamples comprising the total. Correlations between matching factor scores from the total sample and subsample analyses are presented for each factor along with the mean of the correlations among non-matching factors.